A Configurational Bias Monte Carlo Method for Linear and Cyclic Peptides

Michael W. Deem* and Joel S. Bader

CuraGen Corporation

322 East Main Street

Branford, CT 06405

February 1, 2008

Running Title: Biased Monte Carlo of Cyclic Peptides.

*Author to whom correspondence should be addressed. Present address: Lyman Laboratory of Physics, Harvard University, Cambridge, MA 02138.

Abstract

In this manuscript, we describe a new configurational bias Monte Carlo technique for the simulation of peptides. We focus on the biologically relevant cases of linear and cyclic peptides. Our approach leads to an efficient, Boltzmann-weighted sampling of the torsional degrees of freedom in these biological molecules, a feat not possible with previous Monte Carlo and molecular dynamics methods.

1   Introduction

This paper presents a new Monte Carlo method that employs biased trial moves to achieve an efficient sampling of the torsional degrees of freedom for linear and cyclic peptides.

Peptides are small molecules, built from amino acids, that are of fundamental importance in biological systems [1]. They play key roles in signal transduction between cells, regulation of cell growth and differentiation, and protein localization on cell surfaces [2]. Peptides are thought to regulate neurotransmission, from modulating pain and thirst to affecting memory and emotion [3, 4]. They are used as a chemical defense mechanism by some organisms. The *conus* snails, for example, produce a family of highly-constrained peptides that include very powerful neurotoxins [5]. Finally, peptides are used within the biotechnology industry to identify antagonists blocking various abnormal enzymatic actions or ligand-receptor interactions [6]. Cyclic or otherwise constrained peptides are often preferred for this application, since such molecules suffer less of a loss of configurational entropy upon binding [7]. A classic example is the use of the RGD peptide to block the GPIIb/IIIa-fibronectin interaction, reducing blood platelet aggregation [8, 9].

The properties of peptides are amenable to examination by computer experiment. An early study was of the alanine dipeptide, in which the potential energy surface was deduced from *ab initio* quantum mechanical calculations [10, 11]. Larger peptides have been examined by classical simulations. Both molecular dynamics [12] and Monte Carlo [13] approaches have proven useful. The effects of the aqueous environment have been incorporated by simple dielectric theory [14, 15, 16, 17] or by explicit inclusion of water molecules [18].

It has become clear, however, that the standard molecular dynamics and Monte Carlo methods are not capable of sampling all conformational degrees of freedom accessible at body temperature to the larger peptides. This problem is particularly evident for the important case of constrained peptides. Various solutions, such as high-temperature molecular dynamics [19, 20] or simplified force fields [20, 21], have been suggested, but these approaches suffer from uncontrolled approximations. A simulation method able to sample the relevant conformational states of peptides, particularly constrained ones, or exposed loops of larger proteins would be of great value. It would aid study of these molecules in biological systems as well as facilitate structural understanding of the peptides and antibodies of interest to the biotechnology industry.

Recently, powerful Monte Carlo methods have been developed that have a greatly enhanced sampling efficiency [22, 23, 24, 25, 26, 27, 28, 29]. These methods have been applied to chain molecules at low and high density [24, 30] and even at phase coexistence [31, 32, 33, 34]. These methods all use importance sampling, or biased moves, to efficiently explore the free energy landscape.

We here apply these concepts to peptide molecules. Both linear and constrained or cyclic peptides are treated by this method. In Sec. 2 we describe the Monte Carlo method in detail. Appendices describe the rigid molecular fragments from which peptides are constructed and provide technical details of the method. In Sec. 3 we describe the application of this method to the prototypical polyglycine peptides. We discuss the results in Sec. 4. The superiority of this method over conventional molecular dynamics and Monte Carlo is demonstrated. Conclusions are presented in Sec. 5.

2   Monte Carlo Method

We make the simplifying assumption that the intramolecular potential energy contains only torsional and non-bonded terms. That is, bond lengths and angles are fixed, and rotation is allowed only about sigma bonds. At room- or body-temperature, these are fairly good assumptions. They could easily be relaxed, although sampling the increased degrees of freedom would entail a computational expense. Appendix A describes the rigid fragments that occur in peptides under these assumptions. A suitable form for the interatomic potential would be the AMBER [35], ECCEP [36], or CHARMm [37] force field. We pick the AMBER potentials. Water is treated in an implicit way, assuming the dielectric constant for Coulomb interactions is given by $\epsilon/\epsilon_0 = 4r$, with r given in Ångstroms. These assumptions allow the method to be presented without a discussion of detailed force field issues. The method is generically applicable to better force fields and an explicit treatment of water.

A configurational bias Monte Carlo (CBMC) technique is used to explore the conformations of the molecules. We describe the algorithm for both linear and cyclic peptides. By cyclic, we mean peptides constrained because of disulfide bonds between cystine residues.

There are two types of atoms in a peptide, those in the side chains and those in the backbone. Consequently, there are two types of Monte Carlo moves: type I moves change the positions of side chain atoms only, and type II moves change the positions of backbone atoms, rigidly rotating the attached side chains. The type I move is an extension of the

4

chain-molecule CBMC [24, 25] to the structurally more complicated case of peptides. The type I move is applicable to side chains with a free end (*i.e.* all naturally occurring amino acid side chains except for proline). The backbone to which the side chain is attached can be either linear or cyclic. In the cyclic case, the type I move is also used to change the configuration of the free ends of the main chain.

There are two kinds of type II moves for the backbone: type IIa moves for linear peptides and type IIb moves for cyclic peptides. The type IIa move is essentially the same as a type I move. The side-chain residues that are attached to the backbone are rigidly rotated so as to remain properly bonded to the C_α atoms in their new positions. When the peptide is cyclic, we use a type IIb move to change the configuration of part of the backbone loop, rigidly rotating any side chains or free ends of the peptide that are attached to that part of the backbone. The backbone of a cyclic peptide includes the atoms along the main chain as well as the C_β and S atoms of the cystines participating in the disulfide bond. This move requires a concerted rotation of the backbone torsional angles with a rigid rotation of the attached side groups. This concerted rotation of the torsional angles is an extension of the concerted rotation scheme for alkanes [22, 28].

A type I move is initiated by identifying the side chain to be regrown. Not all of the side chain need be regrown, and the first group to regrow is chosen. This feature is helpful for the amino acids with longer side chains, such as lysine. These choices are made randomly. The M rigid units to be regrown are first removed and then added one at a time, starting from the one closest to the backbone. For each addition, the following actions are carried out (see Fig. 1):

1) k values of the torsional angle ϕ_{ij}, $1 \leq j \leq k$ connecting rigid unit i to unit $i-1$ are generated according to the internal potential,

$$p_i^{int}(\phi_{ij}) \propto \exp[-\beta u_i^{int}(\phi_{ij})] \ . \tag{1}$$

The function $u_i^{int}(\phi_{ij})$ is the part of the internal energy that couples unit i to the rest of the molecule (but excluding units $i+1$ to M). The inverse temperature is given by $\beta = 1/k_B T$.

2) One of these is picked with probability

$$p_i^{ext}(\phi_{ij}) = \exp[-\beta u_i^{ext}(\phi_{ij})]/w^{ext}(i) \ , \tag{2}$$

where

$$w^{ext}(i) = \sum_{j=1}^{k} \exp[-\beta u_i^{ext}(\phi_{ij})] \ . \tag{3}$$

5

The function $u_i^{ext}(\phi_{ij})$ is the part of the external energy that couples unit i to the rest of the molecule (but excluding units $i+1$ to M).

3) Steps 1-2 are repeated until all M units have been added.

4) The Rosenbluth weight

$$W^{(n)} = \prod_{i=1}^{M} w^{ext}(i) \tag{4}$$

is calculated. This attempted move is accepted with a probability

$$acc(o \to n) = \min[1, W^{(n)}/W^{(o)}] . \tag{5}$$

The quantity $W^{(o)}$ is the Rosenbluth weight for the reverse move and is calculated as in steps 2-4, but with $k-1$ random orientations and one orientation that is equal to the original geometry for each rigid unit.

A type IIa move is very similar to a type I move. In this case, the direction of regrowth is chosen randomly. Then the first backbone unit to be regrown is chosen. The M rigid units to be regrown are removed and added back sequentially, as in the type I move. The rigid units in this case are either A-units, B-units with the side chain rigidly attached, C-units, or D-units (see appendix A). An alternative procedure would be to regrow the side chain units as well, but this proved not to be efficient, due to frequent steric repulsions. The move is accepted with the probability given by Eq. (5).

A type IIb move is initiated by identifying the 4 rigid units on the backbone to be rotated. This is done randomly. The four rigid units are labeled in an amine to carboxy terminal fashion. The attached side groups are rigidly rotated with the backbone units.

The rotation is carried out as follows (see Fig. 2):

1) The driver angle ϕ_0 is changed by an amount $\delta\phi_0$, where $-\Delta\phi < \delta\phi_0 < \Delta\phi$. This is done k' times with probabilities according to the internal potential,

$$p^{int}(\phi_{0j}) \propto \exp[-\beta u_0^{int}(\phi_{0j})] . \tag{6}$$

The function $u_0^{int}(\phi_{0j})$ is the internal energy associated with this torsional angle. Only those values of ϕ_0 that lead to valid solutions for the modified torsional angles are considered. In the general case there will be a distinct ϕ_1 for each solution arising from the new value of ϕ_0. Define $k^{(n)}$ to be the number of ϕ_0-ϕ_1 pairs. If $k^{(n)} = 0$, the move is rejected.

2) A ϕ_0-ϕ_1 pair is picked with probability

$$p_0^{ext}(\phi_{0j}, \phi_{1j}) = \exp[-\beta u_0^{ext}(\phi_{0j}, \phi_{1j})]/W^{(n)} , \tag{7}$$

where

$$W^{(n)} = \sum_{j=1}^{k^{(n)}} \exp[-\beta u_0^{ext}(\phi_{0j}, \phi_{1j})] \ . \tag{8}$$

The function $u_0^{ext}(\phi_{0j}, \phi_{1j})$ is the part of the external energy that couples this part of the backbone to the rest of the molecule. The value $J^{(n)}$ of the Jacobian is calculated for the new, chosen configuration (as detailed in Appendix B).

3) The reverse move is considered. That is, a rotation about the new, chosen ϕ_0-ϕ_1 pair is considered. $k' - 1$ random values $\delta\phi_0$ are chosen. The original value of ϕ_0 is assigned to the k'th value. This move results in $k^{(o)}$ solutions for ϕ_1. $k^{(o)}$ is always greater than zero, since the original configuration exists. (Special care is taken to ensure that the original configuration is found by the root finding procedure.) The Rosenbluth weight is assigned to $W^{(o)}$. The value $J^{(o)}$ of the Jacobian is also calculated for the original configuration.

This attempted move is accepted with a probability

$$acc(o \to n) = \min[1, J^{(n)}W^{(n)}/J^{(o)}W^{(o)}] \ . \tag{9}$$

Splitting the energy into internal and external parts is rather arbitrary. There are some constraints imposed, however, by the requirement that the normalization constants for Eqs. (1) and (6) be independent of chain conformation [26]. We assume for simplicity that $u_i^{int} = 0$. One other natural choice, however, would set the internal part equal to the torsional terms in H_{intra} and set the external part equal to the rest of H.

For any Monte Carlo scheme to properly sample the Boltzmann probability distribution, detailed balance must be satisfied. Refs. [22] and [26] prove that detailed balance is satisfied for the above scheme.

3   Application to Polyglycine

In this section we present the result of applying this configurational bias Monte Carlo method to two simple peptides, polyglycine G_6 and constrained polyglycine CG_6C.

Figure 3 shows the energy of linear polyglycine as a function of Monte Carlo steps. This run took roughly 3 hours on a Silicon Graphics Indigo^2. In Fig. 4 we show the end-to-end probability distribution for this system. Gaining this degree of convergence took a one-day run.

Figure 5 illustrates the energy of the cyclic polyglycine as a function of Monte Carlo steps. This run took roughly 6 hours. Figure 6 provides a histogram of the number

of solutions found for each attempted concerted rotation. In rare cases the root finding procedure failed to find all the roots. In the construction of this plot, we rounded $n^{(n)}$ up when it was odd. Figure 7 shows the histogram for the $C_\beta SSC_\beta$ dihedral angle, with the statistics taken from a run six times as long as that illustrated in Fig. 5. To give a feel for the barrier to rotation about this angle, we show in Fig. 8 the potential of mean force. This potential was determined by umbrella sampling [38]. This curve took two orders of magnitude longer to determine than did the probability distribution in Fig. 7. The potential of mean force is contrasted with the energy associated purely with the $C_\beta SSC_\beta$ torsional terms. Finally, Fig. 9 shows the result of classifying the configurations produced by the method into distinct stable conformations. Fuzzy clustering [39] was used to determine the dominant conformations, with the result that there are only two or three distinct conformations within this limited simulation run. The simulation run depicted in Figs. 5 and 9 took approximately 8 hours on a Silicon Graphics Indigo^2.

4   Discussion

We see that with a very modest computational effort, we can achieve equilibrated results for linear peptides. With somewhat more effort, we can achieve equilibration for cyclic peptides.

As expected, we find that the linear peptide G_6 is relatively unstructured in solution. There is a common crumpled state, but there is also a significant population of the extended state. The constraint of the disulfide bond in CG_6C, in contrast, forces that molecule to adopt a limited number of molecular conformations. For the fairly short runs illustrated in Figs. 5,6,7 and 9, we find only three dominant conformations. The first conformation is associated with the $C_\beta SSC_\beta$ torsional angle of 290°, whereas the other two are associated with angles of 88° and 98°. The first of these conformations is very tight, with 0.7 Å fluctuations about the mean for all atoms in the molecule. The other two are somewhat looser, with roughly 1.2 Å fluctuations. We see from Fig. 9 that even in this short run the method revisits previous conformations. In the limit of a long simulation, the time spent in each conformation would, of course, be proportional to the exponential of the free energy of the conformation.

If CG_6C were achiral, the potential of mean force in Fig. 8 would be symmetric about 0° and 180°. Since the C_α carbons in the cystine residues are, in fact, chiral, the potential of mean force is not required to be symmetric. The asymmetry seen in Fig. 8 results from

the mean, chiral force of the rest of the molecule on the $C_\beta SSC_\beta$ torsion. In fact, the AMBER forcefield takes this chirality into account by reducing the symmetry of the C_β carbon in cysteine. We have used this geometry [40]. The barrier at $0°$ is due to a high steric repulsion between the hydrogens on the C_β carbons adjacent to the disulfide bond. This barrier is substantially higher than the barrier at $180°$.

From Fig. 8, we see that there is a very significant free energy barrier to rotation about the $C_\beta SSC_\beta$ torsional angle. This figure was not constructed from a standard simulation run, but by the specialized procedure of umbrella sampling. It is clear from Fig. 7, however, that the present method is able to overcome this barrier and to properly sample the relevant conformations even in a relatively short simulation. Any method such as molecular dynamics or standard Monte Carlo that makes only small, local changes to the configuration would never cross this barrier in a simulation of reasonable length. High temperature dynamics can allow systems to cross high barriers, but can not perform the requisite Boltzmann sampling to predict the physiologically relevant conformations. Only a biased method that makes fairly large geometrical changes is capable of dealing with such barriers in an automatic way, without resort to special techniques such as umbrella sampling. Furthermore, the ability to perform umbrella sampling has as a prerequisite the detailed knowledge of the important conformations and the paths between them. In our specific case, we find our method to be two orders of magnitude more efficient than umbrella sampling.

5 Conclusion

We have presented a Monte Carlo method capable of sampling the relevant room- or body-temperature configurations of linear and cyclic peptides. This method allows the study of peptides important in biological and technological settings. Our sampling of the disulfide dihedral angle in a prototypical cyclic peptide indicates that the method can explore widely separated regions of conformation space according to the proper Boltzmann distribution, even if the barriers between the regions are quite large. Previous simulation methods either fail to sample the proper thermal distribution or are vastly more computationally intensive and require detailed knowledge of the thermally accessible regions. The method can be extended to allow incorporation of explicit water molecules. The method can be extended to force fields with flexible bonds and angles. These extensions are subjects for future work.

Acknowledgements

References

[1] Alberts, B., 1994, *Molecular Biology of the Cell*, 3rd edition, (Garland: New York).

[2] Cohen, G. B., Ren, R. B., and Baltimore, D., 1995, *Cell*, **80**, 237.

[3] Kandel, E., and Abel, T., 1995, *Science*, **268**, 825.

[4] Swerdlow, J. L., 1995, *Natl. Geog.*, **187**(6), 2.

[5] Olivera, B. M., *et al.*, 1990, *Science*, **249**, 257.

[6] Clackson, T. and Wells, J. A., 1994, *Trends in Biotechnology*, **12**, 173.

[7] Alberg, D. G., and Schreiber, S. L., 1993, *Science*, **262**, 248.

[8] Ruoslahti, E., 1992, *British Journal of Cancer*, **66**, 239.

[9] O'Neil, K. T., Hoess, R. H., Jackson, S. A., Ramachandran,N. S., Mousa, S. A., and DeGrado, W. F., 1992, *Proteins: Structure, Function, and Genetics*, **14**, 509.

[10] Cheam, T. C., and Krimm, S., 1990, *Theochem Journal of Molecular Structure*, **65**, 173.

[11] Tobias, D. J., and Brooks, C. L., 1992, *J. Phys. Chem.*, **96**, 3864.

[12] Roux, B., and Karplus, M., 1994, *Annual Review of Biophysics and Biomolecular Structure*, **23**, 731.

[13] Nikiforovich, G. V., 1994, *Int. J. Peptide Protein Res.*, **44**, 513.

[14] Schiffer, C. A., Caldwell, J. W., Kollman, P. A., and Stroud, R. M., 1993, *Mol. Simulat.*, **10**, 121.

[15] Smith, P. E., and Pettitt, B. M., 1993, *J. Phys. Chem.*, **97**, 6907.

[16] Gould, I. R., and Hillier, I. H., 1993, *J. Chem. Soc. Chem. Comm.*, **11**, 951.

[17] Daggett, V., Kollman, P. A., and Kuntz, I. D., 1991, *Biopolymers*, **31**, 285.

[18] Yan, Y. B., Tropsha, A., Hermans, J., and Erickson, B. W., 1993, *Proc. Nat. Acad. Sci.*, **90**, 7898.

[19] Bruccoleri, R. E., and Karplus, M., 1990, *Biopolymers*, **29**, 1847.

[20] Tsujishita, H., Moriguchi, I., and Hirono, S., 1994, *Biophysical Journal*, **66**, 1815.

[21] Brunne, R. M., Van Gunsteren, W. F., Bruschweiler, R., and Ernst, R. R., 1993, *J. Am. Chem. Soc.*, **115** 4764.

[22] Dodd, L. R., Boone, T. D., and Theodorou, D. N., 1993, *Molec. Phys.*, **78**, 961.

[23] Frenkel, D., Mooij, G. C. A. M., and Smit, B., 1992, *J. Phys.: Condens. Matter*, **4**, 3053.

[24] Frenkel, D., and Smit, B., 1992, *Molec. Phys.*, **75**, 983.

[25] DePablo, J. J., Laso, M., and Suter, U. W., 1992, *J. Chem. Phys.*, **96**, 6157.

[26] Smit, B., and Siepmann, J. I., 1994, *J. Phys. Chem.*, **98**, 8442.

[27] Maginn, E. J., Bell, A. T., and Theodorou, D. N., 1995, *J. Phys. Chem.*, **99**, 2057.

[28] Leontidis, E., de Pablo, J. J., Laso, M,. and Suter, U. W., 1994, *Adv. Pol. Sci.*, **116**, 283.

[29] Escobedo, F. A., and de Pablo, J. J., 1995, *J. Chem. Phys.*, **102**, 2636.

[30] DePablo, J. J., Laso, M., Siepmann, J. I., and Suter, U. W., 1993, *Molec. Phys.*, **80**, 55.

[31] Mooij, G. C. A. M., Frenkel, D., and Smit, B., 1992, *J. Phys.: Condens. Matter*, **4**, L255.

[32] Laso, M., DePablo, J. J., and Suter, U. W., 1992, *J. Chem. Phys.*, **97**, 2817.

[33] Siepmann, J. I., Karaborni, S., and Smit, B., 1993, *Nature*, **365**, 330.

[34] Smit, B., Karaborni, S., and Siepmann, J. I., 1995, *J. Chem. Phys.*, **102**, 2126.

[35] Weiner, S. J., *et al.*, 1986, *J. Comp. Chem.*, **7**, 230.

[36] Nemethy, G., Pottle, M. S., and Scheraga, H. A., 1983, *J. Phys. Chem.*, **87**, 1883.

[37] Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J. Swaminathan, S. and Karplus, M., 1983, *J. Comput. Chem.*, **4**, 187.

[38] Chandler, D., 1987, *Introduction to Modern Statistical Mechanics*, (Oxford University Press: New York), Ch. 6.3.

[39] Gordon, H. L., and Somorjai, R. L., 1992, *Proteins: Structure, Function, and Genetics*, **14**, 249.

[40] From the InsightII program, version 2.9.5, BIOSYM Technologies, 9685 Scranton Road, San Diego, CA 92121.

[41] Ryckaert, J. P., Ciccotti, G., Berendsen, H. J. C., 1977, *J. Comp. Phys.*, **23**, 327.

[42] Shenkin, P. S., Yarmush, D. L., Fine, R. M., Wang, H. J., and Levinthal, C., 1987, *Biopolymers*, **26**, 2053.

Appendix A: Rigid Units

As described, the algorithm assumes that bond lengths and angles are fixed. The only degrees of freedom, therefore, are torsional angles. Due to the extremely high force constant for rotation about a π bond, even some torsional angles are fixed as well. An entire collection of atoms that is rigid is called a rigid unit. Such a unit has an incoming bond as well as several possible outgoing bonds. There are four backbone rigid units. Unit A is the starting NH_3^+ group. Unit D is the terminal COO^- group. Unit B is the $C_\alpha H$ group. Unit C is the CONH amide bond group.

The residues are connected to the backbone by outgoing bonds from the B units. Table 1 lists the decomposition of the amino acid side chains into rigid units. Typical rigid

Table 1: The rigid units in peptide side groups.

Side Group	Rigid Units
Glycine	H
Alanine	CH_3
Arginine	CH_2, CH_2, CH_2, $CN_3H_5^+$
Aspartate	CH_2, CO_2^-
Asparagine	CH_2, $CONH_2$
Cyst(e)ine	CH_2, $S(H)$
Glutamate	CH_2, CH_2, CO_2^-
Glutamine	CH_2, CH_2, $CONH_2$
Histidine	CH_2, $C_3N_2H_3$
Isoleucine	CH, CH_2, CH_3, CH_3
Leucine	CH_2, CH, CH_3, CH_3
Lysine	CH_2, CH_2, CH_2, CH_2, NH_3^+
Methionine	CH_2, CH_2, S, CH_3
Phenylalanine	CH_2, C_6H_5
Proline	Backbone Groups: $C_\alpha HCH_2CH_2CH_2$, N, CO
Serine	CH_2, OH
Threonine	CH, CH_3, OH
Tryptophan	$CH2$, C_8NH_6
Valine	CH, CH_3, CH_3
Tyrosine	CH_2, C_6H_4, OH

units are the CH_2, CN_3, CO_2, and aromatic ring groups, which have substantial π bonding character.

Proline is a special case, technically an imino acid. The special nature is due to the cyclic bonding of the residue to the backbone. The rigid units in this amino acid are the CH_n, CO, and N groups. Only *trans* isomers are allowed for the proline amide bond. Proline is treated in an approximate way: the C_α-C_δ fragment is kept rigid, the C_δ-N bond is broken, and the C_α-N torsional barrier is increased. This approximation ignores the small fluctuations in the configuration of the proline side-chain loop.

Appendix B: Concerted Rotation

Since the molecules under consideration can be cyclic, a Monte Carlo move that preserves this constraint is required. The "concerted rotation" scheme used for alkanes [22] can be extended to allow rotation of the torsional angles in cyclic peptides. This appendix describes this extension. The reader is referred to Ref. [22] for a fuller discussion of the original, restricted method. The method presented here allows for a fairly general molecular geometry. In particular, the method naturally accommodates the constraint of a planar amide bond.

To formulate the method, we consider rotating about seven torsional angles, which will move the root positions of four rigid units, rotate up to three additional ones, and leave the rest of the peptide fixed. We define the root position of a rigid unit to be the C_α position for a B unit, the C position for a C unit, the C position for a CH_2 unit, and the S position for the S unit in cystine. If unit 5 is a C unit, however, $\mathbf{r}_5$ is defined to be the N position of that unit. For each unit we define θ_i to be the angle between the incoming and outgoing bonds. Thus, $\theta_i = 0$ for a C unit, and $\theta_i \approx 70.5°$ for all others. Figure 1 illustrates the geometry under consideration.

The method leaves the positions $\mathbf{r}_i$ of units $i \leq 0$ or $i \geq 5$ fixed. The torsion ϕ_0 is changed by an amount $\delta\phi_0$. The values of $\phi_i, 1 \leq i \leq 6$, are then determined so that only the positions $\mathbf{r}_i$ of units $1 \leq i \leq 4$ are changed.

The method required several definitions to present the solution for the new torsional angles. Vectors are defined which are the difference in position between unit i and unit $i - 1$, as seen in the coordinate system of unit i:

$$\mathbf{l}_i = \mathbf{r}_i^{(i)} - \mathbf{r}_{i-1}^{(i)} \ . \tag{10}$$

The coordinate system of i is such that the incoming bond is along the $\hat{\mathbf{x}}$ direction. Thus $\mathbf{l}_i = l_i\hat{\mathbf{x}}$ if atom $\mathbf{r}_i$ and $\mathbf{r}_{i-1}$ are directly bonded and has x- and y-components otherwise. We now define a rotation matrix that transforms from the coordinate system of unit $i + 1$ to unit i

$$\mathbf{T}_i = \begin{pmatrix} \cos\theta_i & \sin\theta_i & 0 \\ \sin\theta_i\cos\phi_i & -\cos\theta_i\cos\phi_i & \sin\phi_i \\ \sin\theta_i\sin\phi_i & -\cos\theta_i\sin\phi_i & -\cos\phi_i \end{pmatrix} \ . \tag{11}$$

The positions of the units in the frame of unit 1 are, thus, given by

$$
\begin{aligned}
\mathbf{r}_1^{(1)} &= \mathbf{l}_1 \\
\mathbf{r}_2^{(1)} &= \mathbf{l}_1 + \mathbf{T}_1 \mathbf{l}_2 \\
\mathbf{r}_3^{(1)} &= \mathbf{l}_1 + \mathbf{T}_1 (\mathbf{l}_2 + \mathbf{T}_2 \mathbf{l}_3) \\
\mathbf{r}_4^{(1)} &= \mathbf{l}_1 + \mathbf{T}_1 (\mathbf{l}_2 + \mathbf{T}_2 (\mathbf{l}_3 + \mathbf{T}_3 \mathbf{l}_4)) \ .
\end{aligned}
\tag{12}
$$

We further define the matrix that converts from the frame of reference of unit 1 to the laboratory reference frame

$$
\mathbf{T}_1^{lab} = [\cos \psi \mathbf{I} + \mathbf{n} \mathbf{n}^\top (1 - \cos \psi) + \mathbf{M} \sin \psi] \mathbf{A} \ ,
\tag{13}
$$

where

$$
\mathbf{M} = \begin{pmatrix} 0 & -n_z & n_y \\ n_z & 0 & -n_x \\ -n_y & n_x & 0 \end{pmatrix} ,
\tag{14}
$$

and

$$
\begin{aligned}
\mathbf{n} &= \frac{\hat{\mathbf{x}} \times \mathbf{r}}{|\hat{\mathbf{x}} \times \mathbf{r}|} \\
\cos \psi &= \frac{\mathbf{r} \cdot \hat{\mathbf{x}}}{|\mathbf{r}|} \\
\sin \psi &= \frac{|\mathbf{r} \times \hat{\mathbf{x}}|}{|\mathbf{r}|} \ ,
\end{aligned}
\tag{15}
$$

where $\hat{\mathbf{x}}$ is a laboratory unit vector along the x direction, and $\mathbf{r}$ is the axis of the bond coming into unit 1. The matrix $\mathbf{A}$ is a rotation about $\hat{\mathbf{x}}$ and is defined so that $\mathbf{A}\mathbf{l}_1 = \Delta \mathbf{r}$:

$$
\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c & -s \\ 0 & s & c \end{pmatrix}
\tag{16}
$$

where

$$
\begin{aligned}
c &= (l_{1y} \Delta r_y + l_{1z} \Delta r_z)/(\Delta r_y^2 + \Delta r_z^2) \\
s &= (-l_{1z} \Delta r_y + l_{1y} \Delta r_z)/(\Delta r_y^2 + \Delta r_z^2) \ .
\end{aligned}
\tag{17}
$$

15

Here $\Delta \mathbf{r} = \mathbf{A}[\mathbf{T}_1^{lab}]^{-1}(\mathbf{r}_1 - \mathbf{r}_0)$ if unit 0 is a C unit; otherwise $\Delta \mathbf{r} = \mathbf{l}_1$.

The method proceeds by solving for $\phi_i, 2 \leq i \leq 6$, analytically in terms of ϕ_1. Then a nonlinear equation is solved numerically to determine which values of ϕ_1, if any, are possible for the chosen value of ϕ_0.

We will work in the coordinate system of unit 1, after it has been rotated by the chosen ϕ_0. We define

$$\mathbf{t} = \mathbf{r}_5^{(1)} - \mathbf{l}_1 = [\mathbf{T}_1^{lab}]^{-1}(\mathbf{r}_5 - \mathbf{r}_0) - \mathbf{l}_1 \ . \tag{18}$$

If $\theta_3 \neq 0$ and $\theta_5 \neq 0$, the square distance between unit 3 and unit 5 is known and equal to

$$q_1^2 = (l_{4x} \cos \theta_4 - l_{4y} \sin \theta_4 + l_{5x})^2 + (l_{4x} \sin \theta_4 + l_{4y} \cos \theta_4 + l_{5y})^2 \ . \tag{19}$$

But this distance can also be written as

$$\begin{aligned} q_1^2 &= |\mathbf{x} - \mathbf{T}_2 \mathbf{l}_3|^2 \\ \mathbf{x} &= \mathbf{T}_1^{-1}\mathbf{t} - \mathbf{l}_2 \ . \end{aligned} \tag{20}$$

Equating these two results, two values of ϕ_2 are possible

$$\begin{aligned} \phi_2^{\mathrm{I}} &= \arcsin c_1 - \arctan x_y/x_z - H(x_z) \\ \phi_2^{\mathrm{II}} &= \pi - \arcsin c_1 - \arctan x_y/x_z - H(x_z) \ , \end{aligned} \tag{21}$$

with

$$H(x) = \begin{cases} 0, \ x > 0 \\ \\ \pi, \ x < 0 \end{cases} \ . \tag{22}$$

The constant c_1 is given by

$$c_1 = \begin{cases} \dfrac{q_1^2 - x^2 - l_3^2 + 2x_x(\cos \theta_2 l_{3x} + \sin \theta_2 l_{3y})}{-2(\sin \theta_2 l_{3x} - \cos \theta_2 l_{3y})(x_y^2 + x_z^2)^{1/2}}, \ \theta_3 \neq 0, \theta_5 \neq 0 \\ \\ \dfrac{l_{3x} + l_{4x} + l_{5x} \cos \theta_4 - x_x \cos \theta_2}{\sin \theta_2 (x_y^2 + x_z^2)^{1/2}}, \ \theta_3 = 0, \theta_5 \neq 0 \\ \\ \dfrac{(\mathbf{r}_5 - \mathbf{r}_2) \cdot (\mathbf{r}_6 - \mathbf{r}_5)/l_6 - l_{5x} - l_{4x} \cos \theta_4 - x_x(\cos \theta_2 l_{3x} + \sin \theta_2 l_{3y})}{(\sin \theta_2 l_{3x} - \cos \theta_2 l_{3y})(x_y^2 + x_z^2)^{1/2}}, \ \theta_3 \neq 0, \theta_5 = 0 \\ \\ \dfrac{l_{3x} \cos \theta_4 - x_x(\cos \theta_2 l_{3x} + \sin \theta_2 l_{3y})}{(\sin \theta_2 l_{3x} - \cos \theta_2 l_{3y})(x_y^2 + x_z^2)^{1/2}}, \ \theta_3 = 0, \theta_5 = 0 \end{cases} \ , \tag{23}$$

where $\mathbf{x}$ is given by Eq. (20) if $\theta_5 \neq 0$, and $\mathbf{x} = \mathbf{T}_1^{-1}[\mathbf{T}_1^{lab}]^{-1}(\mathbf{r}_6 - \mathbf{r}_5)/l_6$ if $\theta_5 = 0$. Clearly for there to be a solution $|c_1| \leq 1$. The last three equations for c_1 were determined by

conditions similar to equating Eqs. (19) and (20). For $\theta_3 = 0, \theta_5 \neq 0$, the x-component of $\mathbf{r}_5^{(3)} - \mathbf{r}_3^{(3)}$ is known to be equal to $l_{4x} + l_5 \cos \theta_4$. For $\theta_3 \neq 0, \theta_5 = 0$, the x-component of $\mathbf{r}_5^{(5)} - \mathbf{r}_3^{(5)}$ is known to be equal to $l_{5x} + l_{4x} \cos \theta_4$. For $\theta_3 = 0, \theta_5 = 0$, the angle between $\mathbf{r}_3 - \mathbf{r}_2$ and $\mathbf{r}_6 - \mathbf{r}_5$ is known to be equal to θ_4.

To determine ϕ_3, two expressions for $|\mathbf{r}_5 - \mathbf{r}_4|^2$ are again equated to determine

$$c_2 = \frac{l_5^2 - y^2 - l_4^2 + 2y_x(\cos \theta_3 l_{4x} + \sin \theta_3 l_{4y})}{-2(\sin \theta_3 l_{4x} - \cos \theta_3 l_{4y})(y_y^2 + y_z^2)^{1/2}} \tag{24}$$

and

$$\begin{aligned} \phi_3^{\mathrm{I}} &= \arcsin c_2 - \arctan y_y/y_z - H(y_z) \\ \phi_3^{\mathrm{II}} &= \pi - \arcsin c_2 - \arctan y_y/y_z - H(y_z) , \end{aligned} \tag{25}$$

where $\mathbf{y} = \mathbf{T}_2^{-1}(\mathbf{T}_1^{-1}\mathbf{t} - \mathbf{l}_2) - \mathbf{l}_3$. Again, $|c_2| \leq 1$ for there to be a solution.

If $\theta_5 \neq 0$, the value of ϕ_4 can be determined from

$$\mathbf{r}_5^{(1)} = \mathbf{r}_4^{(1)} + \mathbf{T}_1 \mathbf{T}_2 \mathbf{T}_3 \mathbf{T}_4 \mathbf{l}_5 . \tag{26}$$

Defining

$$\mathbf{q}_3 = \mathbf{T}_3^{-1}\mathbf{T}_2^{-1}\mathbf{T}_1^{-1}[\mathbf{T}_1^{lab}]^{-1}(\mathbf{r}_5 - \mathbf{r}_4) , \tag{27}$$

the equations that define ϕ_4 are given by

$$\begin{aligned} q_{3y} &= \cos \phi_4 (\sin \theta_4 l_{5x} - \cos \theta_4 l_{5y}) \\ q_{3z} &= \sin \phi_4 (\sin \theta_4 l_{5x} - \cos \theta_4 l_{5y}) . \end{aligned} \tag{28}$$

This is a successful rotation if the position of $\mathbf{r}_6$ is successfully predicted. That is, the equation

$$\mathbf{r}_6^{(1)} - \mathbf{r}_5^{(1)} = \mathbf{T}_1 \mathbf{T}_2 \mathbf{T}_3 \mathbf{T}_4 \mathbf{T}_5 \mathbf{l}_6 = [\mathbf{T}_1^{lab}]^{-1}(\mathbf{r}_6 - \mathbf{r}_5) \tag{29}$$

must be satisfied. We consider the x-component which implies

$$F(\phi_1) = \begin{cases} (\mathbf{r}_6^{(1)} - \mathbf{r}_5^{(1)})^\top \mathbf{T}_1 \mathbf{T}_2 \mathbf{T}_3 \mathbf{T}_4 \hat{\mathbf{x}} - (l_{6x} \cos \theta_5 + l_{6y} \sin \theta_5) = 0, \ \theta_5 \neq 0 \\[2ex] (\mathbf{r}_4 - \mathbf{r}_3) \cdot (\mathbf{r}_6 - \mathbf{r}_5) - l_4 l_6 \cos \theta_4 = 0, \ \theta_3 \neq 0, \theta_5 = 0 \\[2ex] |\mathbf{r}_6 - \mathbf{r}_4| - \left[(l_{6x} + l_{5x})^2 + l_{5y}^2\right]^{1/2} = 0, \ \theta_3 = 0, \theta_5 = 0 \end{cases} \tag{30}$$

must be satisfied if the rotation is successful. The equations for the case $\theta_5 = 0$ clearly express the geometric conditions required for a successful rotation.

Eq. (30) is the nonlinear equation for ϕ_1 that must be solved. The equation depends only on ϕ_1 because ϕ_2, ϕ_3, and ϕ_4 are determined by Eqs. (21), (25), and (28) in terms of ϕ_1. This equation has between zero and four values for each value of ϕ_1, however, due to the multiple root character of Eqs. (21) and (25). Equation (30) is solved by searching the region $-\pi < \phi < \pi$ for zero crossings. The search is in increments of $\approx 0.04°$. These roots are then refined by a bisection method. There is always an even number of roots, due to the periodic nature of Eq. 30.

The root positions, $\mathbf{r}_i$, are enough to determine the position and orientation of the seven rigid units that are modified by the concerted rotation. Rigid unit 0 is translated so that its root position is at $\mathbf{r}_0$. It is oriented so that its incoming bond vector is along the outgoing bond vector of rigid unit -1. It is then rotated so that its outgoing bond vector ends at $\mathbf{r}_1$. This process is repeated sequentially for rigid units 1 to 6.

Repeated application of the concerted rotation leads to a slightly imperfect structure, due to numerical precision errors. In a practical application, the geometry would be restored to an ideal state by application of the SHAKE [41] or Random Tweak algorithm [42].

The transformation from ϕ_i, $0 \leq i \leq 6$, to the new solution which is constrained to change only $\mathbf{r}_i$, $1 \leq i \leq 4$, actually implies a change in volume element in torsional angle space. This change in volume element is the reason for the appearance of the Jacobian in the acceptance probability. The Jacobian of the transformation for alkanes is calculated in Ref. [22]. It is slightly different here since root position $\mathbf{r}_5$ is not necessarily the head position. The Jacobian is given by

$$J = 1/|\det \mathbf{B}| \, , \tag{31}$$

where the 5×5 matrix B_{ij} is given by the ith component of $\mathbf{u}_j \times (\mathbf{r}_5 - \mathbf{h}_j)$ for $i \leq 3$ and by the $(i-3)$th component of $\mathbf{u}_j \times (\mathbf{r}_6 - \mathbf{r}_5)/|\mathbf{r}_6 - \mathbf{r}_5|$ for $i = 4, 5$. Here $\mathbf{h}_i = \mathbf{r}_i$ except that $\mathbf{h}_5$ is the head position even if $\theta_5 = 0$, and $\mathbf{u}_i$ is the incoming unit bond vector for unit i.

Figure Captions

Figure 1. The type I move applied to the serine side chain.

Figure 2. The type IIb move is illustrated for the case where unit 0 is (a) a B-unit and (b) a C-unit. In each case, the original geometry and the four possible new geometries for the chosen driver angle are shown. In case (a), one of the new geometries is very different

from the original and the other three new ones. The move is shown for a linear peptide, although it is used only on cyclic peptides.

Figure 3. The energy of G_6 as a function of Monte Carlo steps. Note the rapid equilibration.

Figure 4. The probability distribution for the end-to-end distance for G_6. The distance is between the terminal C_α groups.

Figure 5. The energy of CG_6C as a function of Monte Carlo steps. Note the rapid equilibration.

Figure 6. The number of new solutions found for each attempted concerted rotation for CG_6C.

Figure 7. The observed probability distribution for the $C_\beta SSC_\beta$ torsional angle in CG_6C is shown.

Figure 8. The potential of mean force calculated by umbrella sampling for the $C_\beta SSC_\beta$ torsional angle in CG_6C (dashed line). The potential of mean force implied by Fig. 7 is indicated by the solid line. Also shown is the bare torsional energy contribution for this rotation (dotted line).

Figure 9. Shown are the occupation numbers of the configuration in each of the three dominant conformations as a function of Monte Carlo steps (a). Also shown is the all-atom root-mean-square displacement of the configuration from each of the three dominant conformations (b). The curves for conformation 1 are solid, those for 2 are dashed, and those for 3 are short-dashed.

Cluster Membership

RMS from Cluster Center / Å

Monte Carlo steps / 1000